Integrin–collagen complex: a metal–glutamate handshake

Jordi Bella[1] and Helen M Berman[2]*

The recently determined crystal structure of the complex between an integrin I domain and a synthetic collagen peptide shows a collagen triple-helix engaged in specific macromolecular recognition and binding. This structure confirms a previously proposed binding mechanism for integrin I domains and has important implications for integrin activation and signalling.

Addresses: [1]Wellcome Trust Centre for Cell-Matrix Research, School of Biological Sciences, University of Manchester, 2.205 Stopford Building, Oxford Road, Manchester M13 9PT, UK and [2]Department of Chemistry, Rutgers University, 610 Taylor Road, Piscataway, NJ 08854, USA.

*Corresponding author.
E-mail: berman@rcsb.rutgers.edu

Structure 2000, **8**:R121–R126

0969-2126/00/$ – see front matter

Introduction

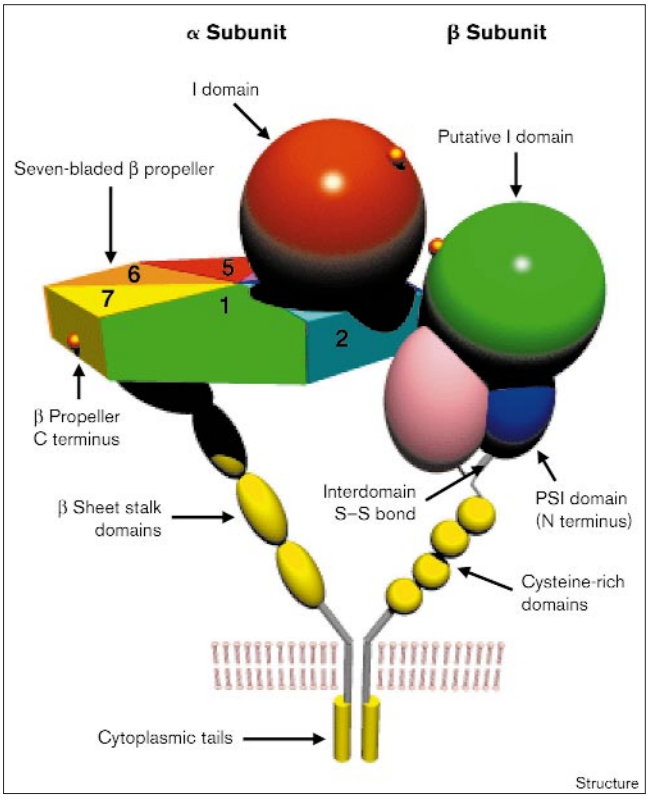
In the March 31st issue of *Cell*, Emsley *et al.* describe the crystal structure of the complex between an integrin fragment and a synthetic collagen peptide that mimics the natural ligand of the integrin [1]. This remarkable structure is a landmark in our understanding of cell–matrix interactions. It provides the first high-resolution view of the interaction between an integrin and one of its ligands. Also for the first time, collagen is shown in great detail acting as a ligand for another protein molecule. Conformational changes seen in the bound integrin fragment may represent the first glimpse of the molecular mechanisms responsible for integrin activation and signal transduction.

Integrins are central molecules in the adhesion processes that mediate cell–cell and cell–extracellular matrix communication. They are heterodimeric cell-surface glycoproteins, composed of two transmembrane polypeptide chains referred to as the α and β subunits. Up to 17 different α subunits and eight β subunits have been identified to date, with 24 different dimers characterised as biologically functional. Integrin ligands have quite varied architectures; they can be cell surface receptors in other cells, plasma proteins, or extracellular matrix proteins such as fibrillar collagen. In addition, some integrins are also subverted as receptors by microbial pathogens. Most integrins are able to bind different ligands with different affinities. The ligand-binding activity of integrins is allosterically controlled by binding events, usually concomitant with signal transmission across the cellular membrane and a subsequent cellular response. This signalling is bidirectional. The interaction of cytoplasmic proteins with integrin cytoplasmic tails can trigger changes in ligand-binding affinity ('inside-out' signalling). Conversely, ligand-binding to the extracellular domains can transduce signals that result in an intracellular response ('outside-in' signalling). Integrin functions relate to many medically relevant processes such as thrombosis, inflammation and tumour metastasis; therefore, the study of integrin–ligand interactions and the development of blocking agents has great therapeutic potential.

The determination of the three-dimensional structure of an integrin dimer is of key importance to the field of adhesion studies, as it would provide critical information on integrin–ligand binding and integrin activation. However, no high-resolution structure of an intact integrin is yet available. Any attempt to crystallise an intact integrin must overcome the fact that both integrin subunits are large, glycosylated, conformationally flexible, and contain membrane-spanning regions. In addition, it is difficult to isolate integrins in large amounts. In the absence of a three-dimensional structure, considerable effort has been spent trying to generate a working model against which biochemical results can be analysed (see [2] for a review). Figure 1 represents one such model of integrin subunit organisation. The N-terminal regions of all integrin α subunits contain seven tandem repeats with internal homology. These repeats have been proposed to fold together in a seven-bladed β-propeller structure [3]. In addition, some integrin α subunits contain an inserted segment of about 200 amino acids in between the second and third repeats. This so-called 'inserted' domain, or I domain, can be expressed in recombinant form isolated from the rest of the integrin. To date, the I domain is the only integrin fragment that has been successfully studied by X-ray crystallography. In the crystal structures reported so far [1,4–7], I domains adopt a Rossmann-type fold with a central set of five parallel β strands and one antiparallel β strand, surrounded by seven α helices (Figure 2). The N and C termini come together at the base of the globular domain, in the correct disposition to connect to the rest of the α subunit.

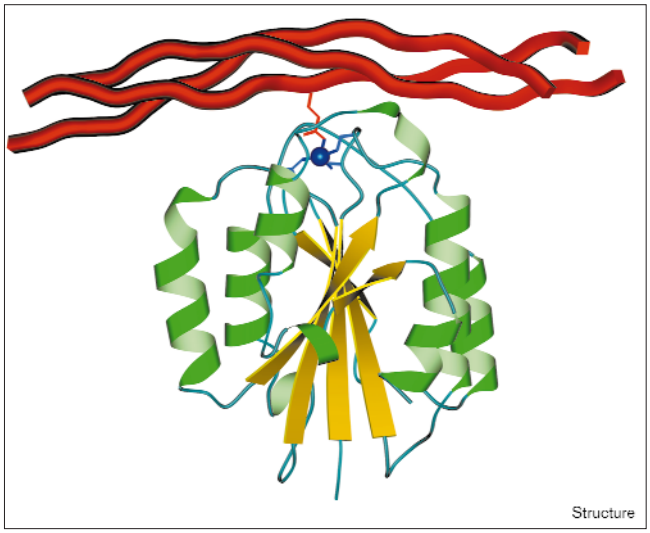
Integrin I domains contain a divalent cation binding site at the top of the five β strands. The metal ion is coordinated by five amino acid sidechains, either directly or indirectly through a water molecule. These residues are located in three loops on the upper surface of the I domain. Three such residues define a conserved signature sequence, DXSXS (in single-letter amino acid code). A highly conserved region of about 200 amino

Figure 1



Schematic representation of a generic integrin heterodimer. Most of the domain structure assignments are predictions made on the basis of sequence homology. The seven N-terminal repeats of the α subunit are represented as a seven-sided disk. When present, the α I domain (red sphere) is inserted between repeats 2 and 3. The N-terminal domain of the β subunit (blue ellipsoid) has been referred to as a plexin-sempahorin-integrin-like (PSI) domain. The green sphere represents the putative β I domain, and the pink ellipsoid represents a region with no known homology. Small orange spheres represent known or proposed divalent cation binding sites. The structures of the cytoplasmic regions are not known.

Figure 2



Ribbon diagram of the α2 I domain in complex with the collagen peptide [1]. Four main types of secondary structure are seen for the first time together in a crystal structure: α helices (green), β strands (yellow), coils (cyan) and collagen triple-helix (red). The metal ion is shown as a blue sphere coordinated to the top of the I domain by three residues, shown in blue, and two water molecules (not shown). A glutamate sidechain from the collagen peptide (in red) completes the coordination of the metal ion. (The figure was produced using the program MOLSCRIPT [25].)

acids in integrin β subunits also contains the DXSXS motif and has potential candidates for the other cation coordinating sidechains. These and other observations suggest that the putative β I domain region adopts a structure related to the α subunit I domain, although not necessarily identical [4,8].

Strong evidence has accumulated over the years for a role of I domains in integrin–ligand interactions. Divalent cations are critical for integrin interactions with almost all ligands, and mutation of the amino acids that define the cation-binding site prevent ligand binding. In the first structure of an integrin I domain [4], the sixth coordination site of the bound magnesium ion was provided by a glutamate sidechain from a neighbouring I domain in the crystal lattice. This serendipitous finding enabled Liddington and colleagues to propose that the symmetry-related molecule was acting as a 'ligand-mimic', and that physiological interactions between integrin I domains and their ligands would involve completing the magnesium coordination with an acidic ligand residue. In support of this hypothesis, I domain integrin ligands, like intercellular adhesion molecule 1 (ICAM-1), have critical glutamic acid residues that when mutated completely abolish binding [9]. Modelling exercises have shown that it is possible to build sterically plausible models of complexes between ICAM-1 and the I domain of its integrin receptor αLβ2, where the critical glutamate residue completes the magnesium coordination [10].

Further I domain crystal structures did not exhibit this type of crystal contact, and most importantly showed significant conformational differences with respect to the first structure. Liddington and colleagues proposed that these conformational changes were significant [11]. The conformation observed in the majority of the crystal structures (with no ligand-mimic contact) was proposed as the 'unliganded' form, whereas the one observed in the structure with the ligand-mimic contact was proposed as the 'liganded' form. Affinity regulation would therefore occur via changes in the coordination in the metal-binding site that were linked to tertiary changes in the entire I domain [11]. This proposal was not universally accepted, and some controversy arose as different groups produced apparently conflicting results, with much of the debate

being published in the pages of this journal [11–14]. As no cation effects were observed in other crystal structures [12,13], the two-conformation model for integrin I domains was called into question.

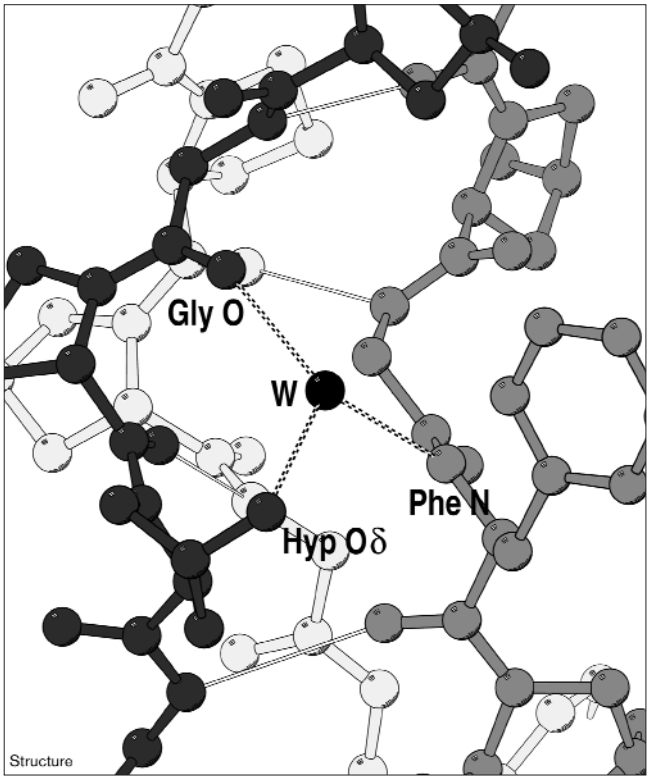
The crystal structure of a complex between an integrin I domain and a fragment of a natural ligand goes a long way towards resolving this controversy [1]. This complex, between the I domain of the α2β1 integrin, known to bind collagen, and a triple-helical collagen peptide containing a critical GFOGER motif (where O stands for hydroxyproline), shows unequivocally how a glutamate residue from the collagen peptide completes the coordination sphere of the I domain metal ion (Figure 2). Most remarkably, comparison with the unliganded α2β1 I domain [6] shows identical conformational changes to those observed in the structure of the αMβ2 I domain involved in a ligand-mimic lattice contact. The fact that these two very different situations produce the same response strongly advocates the correctness of the model initially proposed by Liddington's group in 1995, and suggests that similar changes are to be expected in the conformation of the αLβ2 I domain, for which only the unliganded structure is known [5,12].

A collagen triple-helix engaged in a macromolecular complex

This structure represents the first example of a collagen triple-helix acting as a ligand of another protein molecule. A specific collagen motif has been recently identified as the hexapeptide GFOGER [15]. A peptide containing this sequence was designed and synthesised, and a stable complex was obtained at 4°C. The peptide adopts a triple-helical conformation in the complex, although it bends slightly compared with a standard triple-helical molecule. This bend does not affect the interaction between the recognition sequence and the I domain, and is proposed to be the result of crystal contacts.

The triple-helical conformation is kept through the length of the entire peptide (although slightly distorted at both ends), and keeps the essential characteristics that have been defined in structural studies of peptides with triple-helical conformations [16–19]. The hydrogen-bonding pattern known as Rich and Crick II [20] is kept throughout the entire peptide, and additional water-mediated hydrogen bonds appear between the free amide groups from non-imino acid residues in positions Phe8 and Glu11, and the carbonyl groups in positions Gly4, Gly7 and Gly10 (Figure 3). Such ζ1 water bridges [17] have been recently demonstrated by the crystal structure of T3-785, a triple-helical peptide with a nine-residue track of non-imino acid residues [19], and confirms once again the water-bonded structure proposed by Ramachandran and Chandrasekharan [21]. In the T3-785 crystal structure, water-bridging molecules make additional hydrogen bonds with sidechains from neighbouring hydroxyproline and threonine residues. The same kind of hydrogen bonds are observed in the collagen fragment in complex with the I domain (Figure 3): hydroxyl groups from hydroxyproline residues two positions C-terminal to the glycine residues connected by the water bridge, add extra hydrogen-bonding coordination to the bridging water molecules. These ζ1–γ1 bridges [17] were previously suggested by Ramachandran *et al.* in 1973 [22] as one of the possible roles for hydroxyproline residues in collagen triple-helices. The fact that these water bridges are preserved even in a collagen triple-helix involved in macromolecular contacts strongly suggests that they form an intrinsic part of the collagen triple-helical structure.

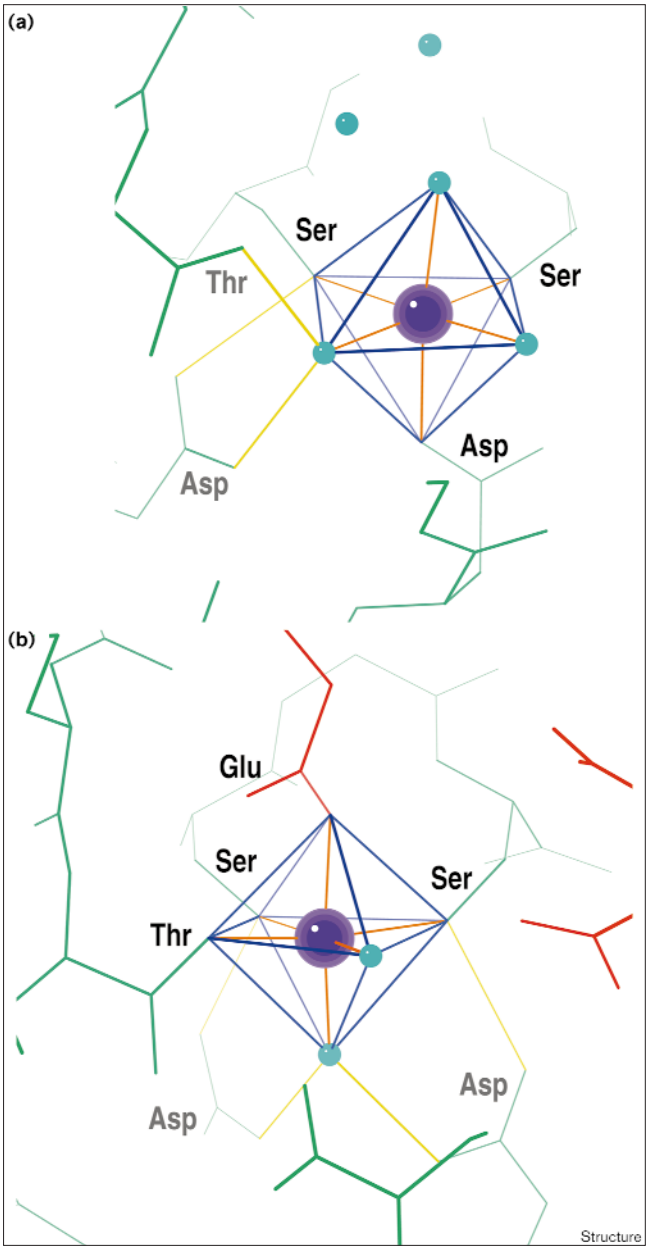
Figure 3



Detail of hydrogen bonding in the collagen peptide from the complex structure of Emsley *et al.* [1]. Each collagen chain is shown in a different shade of grey, and one water molecule (W) is shown as a black sphere. Normal interchain hydrogen bonds are shown as thin white lines. Additional water-mediated hydrogen bonds [19,21,22] (dashed lines) connect the N-H group from a phenylalanine residue to the C=O and hydroxyl groups from a Gly-Pro-Hyp triplet. (The figure was produced using the program MOLSCRIPT [25].)

Remarkably, the helical parameters of the collagen triple-helix in complex with the I domain follow a similar pattern to that observed in the T3-785 crystal structure [19]. The central GFOGER sequence approximates a 10_7 helical symmetry whereas the flanking regions, rich in the imino

Figure 4



Conformational changes associated with changes in metal ion coordination. **(a)** Octahedral coordination of Mg^{2+} or Mn^{2+} (purple sphere) in the closed or unliganded conformation of an I domain. Residues labelled in black are directly coordinated to the metal ion. Water molecules (cyan spheres) complete the coordination. A threonine residue and an aspartate residue (labelled in grey) do not coordinate directly to the metal ion in this conformation but through hydrogen bonds (yellow lines) to water or serine ligands. **(b)** Octahedral coordination of Mg^{2+} or Mn^{2+} (purple sphere) in the open or liganded conformation of an I domain. Sidechains in red correspond to the ligand molecule. Residues labelled in black are directly coordinated to the metal ion, including the threonine residue that was indirectly coordinated in the closed form. An aspartate residue (labelled in grey) that coordinates directly to the metal ion in the closed form does so indirectly through a water molecule in the open form. Yellow lines represent hydrogen bonds.

acids proline and hydroxyproline, approximate a 7_5 symmetry. This reconciles the conflicting observations of helical symmetry between collagen native fibres (tenfold symmetry) and crystal structures of synthetic peptides (sevenfold symmetry). It also confirms an earlier proposal that the high content of imino acids dictates the more twisted, sevenfold conformation in the Pro-Hyp-Gly tracts in the synthetic peptides, whereas a more relaxed tenfold symmetry is expected in collagen molecules with natural, not so imino-acid-rich sequences [16].

Other features of the collagen triple-helical structure, such as δ-type water bridges [17] and Cα-H•••O=C hydrogen bonds [18], are also present in the crystal structure of the collagen fragment in complex with the α2β1 I domain.

The key role of the divalent cation

The structure of the complex formed between the α2β1 I domain and a collagen peptide shows the I domain in a liganded conformation similar to that seen for the Mg^{2+}-bound structure of the αMβ2 I domain [4]. In this structure, a symmetry-related I-domain molecule acted as a ligand mimic. The interacting glutamate residue was part of the C-terminal helix of the I domain, and therefore did not resemble the conformation of the natural ligand for the αMβ2 I domain which is ICAM-1, a molecule with mainly β sheet, immunoglobulin-like structure [10,23]. The collagen triple-helix in complex with the α2β1 I domain resembles neither an α helix nor an immunoglobulin domain. This seems to suggest the main requirement for I domains to go from an unliganded to a liganded conformation is to bind an extra acidic residue, probably glutamate. In principle, any glutamate-containing molecule would suffice, although most probably some affinity threshold must be met in order to produce the conformational change. The many known I domain crystal structures show that the choice of cation is probably irrelevant and similar liganded or unliganded structures are expected for Mg^{2+}, Mn^{2+} or other divalent cations. Ca^{2+}, a notable exception, does not bind with enough affinity to the I domain cation-binding site [13], as it cannot achieve the high coordination numbers characteristic of Ca^{2+}-binding sites.

The different affinities of Mg^{2+} or Mn^{2+} and Ca^{2+} for I domain metal-binding sites may provide some clues as to the molecular mechanisms underlying the conformational change between the liganded and unliganded forms. Figure 4a shows the coordination of the metal centre in the Mn^{2+}-bound αMβ2 I domain (unliganded form). Three residues are directly coordinated to the metal, and the other positions are occupied by water molecules to complete an octahedron. Only one negatively charged residue makes direct contact with the metal (an aspartate), although a second aspartate coordinates indirectly through a water molecule. What happens when an external carboxylic acid comes into the binding position? Figure 4b

shows the metal coordination in the liganded form of the I domain [1,4]. There are three main changes in the liganded state: an aspartate residue that is directly bound to the metal in the unliganded form becomes indirectly bound through a water molecule; one of the water-occupied positions is now coordinated to a threonine residue; and the other water position becomes occupied by the ligand glutamate. In both cases only one carboxylic group is directly bound to the metal. It appears that the metal-binding sites in the I domain structures are unable to hold simultaneously two acidic residues directly coordinated to the metal ion. Therefore, an approaching carboxylic group from a ligand (or ligand mimic) will induce a rearrangement of the coordination around the metal site, and this rearrangement can only be achieved through, or it is responsible for, additional conformational changes that extend across the whole I domain. Interestingly, residues defining the DXSXS motif keep the same coordination (direct or indirect) to the metal ion in both the liganded and unliganded forms.

The way in which subtle changes in metal coordination can amplify to extensive secondary and tertiary changes in the overall I-domain structure is described in detail both by Emsley *et al*. [1] and Lee *et al*. [11]. The main consequence of such changes in the I-domain–collagen complex is the creation of a complementary surface for binding collagen [1]. In addition to the identical metal coordination observed in both liganded forms [1,4], the conformational changes throughout both I domains are very similar, and presumably the liganded form of the αMβ2 I domain represents a 'high affinity' form for its natural ligand, ICAM-1. A similar situation can be expected for the I domain of the αLβ2 integrin [5,12]. It would seem that both the Mg^{2+}- and the Mn^{2+}-bound structures of the αLβ2 I domain (which are essentially identical) correspond to the unliganded form, and that the structure of a complex with ICAM-1 is needed to visualise the expected conformational changes in the liganded form. These changes will probably resemble those seen for the αMβ2 and α2β1 I domains.

I-domain conformation and integrin signalling

The convergence of I-domain structures into essentially two forms suggests that there are just two principal conformations for an I domain: 'open' or liganded, and 'closed' or unliganded [1]. This convergence suggests that it is the formation of appropriate metal–ligand bonds that triggers the conformational transition, although the balance between an open or closed conformation could be shifted by external factors, such as mutations or antibody binding that could stabilise the open conformation. Conformational changes in the I domain will probably transmit across the entire integrin and transduce into a signal across the membrane. How exactly this is achieved is not currently known, but several models have been proposed including some that could account for both outside-in or inside-out signalling [1]. Most probably both the α and β subunits are required for effective signalling. New experimental data for integrins that do not contain an I domain place the second and third blades of the proposed β-propeller in spatial proximity to the putative I-like-domain in the β subunit [24]. As α subunit I domains are inserted between the same second and third blades, they are also likely to be close to the β subunit in a manner similar to the cartoon depicted in Figure 1. Any conformational change in an α subunit I domain is likely to propagate both through its base, across the β-propeller, and sideways, through the β subunit.

Integrins that do not contain α subunit I domains may still bind to ligands through metal-binding sites in the putative I-like-domains in the β subunits. It is known that such regions harbour the binding site for matrix proteins containing RGD or LDV motifs. It is almost impossible to avoid speculating that the aspartate sidechain in such motifs will coordinate a metal ion in a similar way to the glutamate residue in collagen or ICAM-1. If that is the case, there is every possibility that similar changes in metal coordination will trigger changes between two integrin states that will ultimately produce a signalling event.

Acknowledgements

We would like to thank Martin J Humphries, Rachel Kramer and Christine Zardecki for reading of this manuscript, and to Adam Kline and Alex Coe for generating Figure 1. We also want to thank the authors of [1] for the crystal structure coordinates of the I-domain–collagen complex.

References

1. Emsley, J., Knight, C.G., Farndale, R.W., Barnes, M.J. & Liddington, R.C. (2000). Structural basis of collagen recognition by integrin α2β1. *Cell* **100**, 47-56.
2. Humphries, M.J. (2000). Integrin structure. *Biochem. Soc. Trans*., in press.
3. Springer, T.A. (1997). Folding of the N-terminal, ligand-binding region of integrin α-subunits into a β-propeller domain. *Proc. Natl Acad. Sci. USA* **94**, 65-72.
4. Lee, J.-O., Rieu, P., Arnaout, M.A. & Liddington, R.C. (1995). Crystal structure of the A-domain from the α subunit of integrin CR3 (CD11b/CD18). *Cell* **80**, 631-638.
5. Qu, A. & Leahy, D.J. (1995). Crystal structure of the I-domain from the CD11a/CD18 (LFA-1, αLβ2) integrin. *Proc. Natl Acad. Sci. USA* **92**, 10277-10281.
6. Emsley, J., King, S., Bergelson, J. & Liddington, R.C. (1997). Crystal structure of the I domain from integrin α2β1. *J. Biol. Chem.* **272**, 28512-28517.
7. Nolte, M., Pepinsky, R.B., Venyaminov, S.Y., Koteliansky, V., Gotwals, P.J. & Karpusas, M. (1999). Crystal structure of the α1β1 integrin I-domain: insights into integrin I-domain function. *FEBS Lett.* **452**, 379-385.
8. Tuckwell, D.S. & Humphries, M.J. (1997). A structure prediction for the ligand-binding region of the integrin β subunit: evidence for the presence of a von Willebrandt factor A domain. *FEBS Lett.* **400**, 297-303.
9. Humphries, M.J. (1990). The molecular basis and specificity of integrin–ligand interactions. *J. Cell Sci.* **97**, 585-592.
10. Bella, J. Kolatkar, P.R., Marlor, C.W., Greve, J.M. & Rossmann, M.G. (1998). The structure of the two amino-terminal domains of human ICAM-1 suggests how it functions as a rhinovirus receptor and as an LFA-1 integrin ligand. *Proc. Natl Acad. Sci. USA* **95**, 4140-4145.
11. Lee, J.-O., Bankston, L.A., Arnaout, M.A. & Liddington, R.C. (1995). Two conformations of the integrin A-domain (I-domain): a pathway for activation? *Structure* **3**, 1333-1340.
12. Qu, A. & Leahy, D.J. (1996). The role of the divalent cation in the structure of the I domain from the CD11a/CD18 integrin. *Structure* **4**, 931-942.

13. Baldwin, E.T., *et al*., & Wiley, V.H. (1998). Cation binding to the integrin CD11b I domain and activation model assessment. *Structure* **6**, 923-935.
14. Liddington, R.C. & Bankston, L.A. (1998). The integrin I domain: crystals, metals and related artefacts. *Structure* **6**, 937-938.
15. Knight, C.G., Morton, L.F., Peachey, A.R., Tuckwell, D.S., Farndale, R.W. & Barnes, M.J. (2000). The collagen-binding A-domains of integrins α1β1 and α2β1 recognize the same specific amino acid sequence, GFOGER, in native (triple-helical) collagens. *J. Biol. Chem*. **275**, 35-40.
16. Bella, J., Eaton, M., Brodsky, B. & Berman, H.M. (1994). Crystal and molecular structure of a collagen-like peptide at 1.9 Å resolution. *Science* **266**, 75-81.
17. Bella, J., Brodsky, B. & Berman, H.M. (1995). Hydration structure of a collagen peptide. *Structure* **3**, 893-906.
18. Bella, J. & Berman, H.M. (1996). Crystallographic evidence for Cα-H•••O=C hydrogen bonds in a collagen triple helix. *J. Mol. Biol*. **264**, 734-742.
19. Kramer, R.Z., Bella, J., Mayville, P., Brodsky, B. & Berman, H.M. (1999). Sequence dependent conformational variations of collagen triple-helical structure. *Nat. Struct. Biol*. **6**, 454-457.
20. Rich, A. & Crick, F.H.C. (1961). The molecular structure of collagen. *J. Mol. Biol*. **3**, 483-506.
21. Ramachandran, G.N. & Chandrasekharan, R. (1968). Interchain hydrogen bonds via bound water molecules in the collagen triple helix. *Biopolymers* **6**, 1649-1658.
22. Ramachandran, G.N., Bansal, M. & Bhatnagar, R.S. (1973). A hypothesis on the role of hydroxyproline in stabilizing collagen structure. *Biochim. Biophys. Acta* **322**, 166-171.
23. Casasnovas, J.M., Stehle, T., Liu, J.-H., Wang, J.-H. & Springer, T.A. (1998). A dimeric crystal structure for the N-terminal two domains of intercellular adhesion molecule-1. *Proc. Natl Acad. Sci. USA* **95**, 4134-4139.
24. Mould, A.P., Askari, J.A. & Humphries, M.J. (2000). Molecular basis of ligand recognition by integrin α5β1. I. Specificity of ligand binding is determined by amino acid sequences in the second and third NH_2-terminal repeats of the α subunit. *J. Biol. Chem*. April 11 (electronic publication ahead of printed version).
25. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr*. **24**, 946-950.